

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): EBATA, et al.
Serial No.: Not yet assigned
Filed: August 25, 2003
Title: METHOD FOR REBALANCING FREE DISK SPACE AMONG
NETWORK STORAGES VIRTUALIZED INTO A SINGLE FILE
SYSTEM VIEW
Group: Not yet assigned

LETTER CLAIMING RIGHT OF PRIORITY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

August 25, 2003

Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s)
hereby claim(s) the right of priority based on Japanese Patent Application No.(s)
2002-252345, filed August 30, 2002.

A certified copy of said Japanese Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP



Carl I. Brundidge
Registration No. 29,621

CIB/alb
Attachment
(703) 312-6600

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 2002年 8月30日
Date of Application:

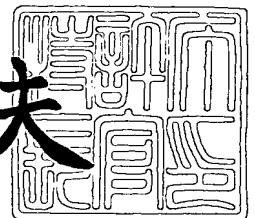
出願番号 特願2002-252345
Application Number:
[ST. 10/C]: [JP 2002-252345]

出願人 株式会社日立製作所
Applicant(s):

2003年 8月14日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫



出証番号 出証特2003-3065695

【書類名】 特許願

【整理番号】 H02010221A

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 12/00

【発明者】

【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目 2 8 0 番地 株式会社日立製作所中央研究所内

【氏名】 江端 淳

【発明者】

【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目 2 8 0 番地 株式会社日立製作所中央研究所内

【氏名】 川本 真一

【発明者】

【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目 2 8 0 番地 株式会社日立製作所中央研究所内

【氏名】 沖津 潤

【発明者】

【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目 2 8 0 番地 株式会社日立製作所中央研究所内

【氏名】 保田 淑子

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社 日立製作所

【代理人】

【識別番号】 100075096

【弁理士】

【氏名又は名称】 作田 康夫

【電話番号】 03-3212-1111

【手数料の表示】

【予納台帳番号】 013088

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 仮想一元化されたネットワークストレージシステムのディスク使用量の平準化方法

【特許請求の範囲】

【請求項 1】

ネットワークに接続された複数のネットワークストレージ装置から構成され、ファイルの格納先ネットワークストレージ装置を管理する配置情報を持ち、該複数のネットワークストレージ装置がクライアントからみて仮想的に一つのネットワークストレージ装置として見える仮想一元化ネットワークストレージシステムにおいて、

前記ネットワークストレージ装置間のデータ移動を伴う仮想一元化ネットワークストレージシステムのディスク残量の平準化方法であって、

各ネットワークストレージ装置のディスク残量を調査するディスク残量調査ステップと、

前記ディスク残量調査ステップの結果から前記ディスク残量の最大値と最小値を求め、該最大値と最小値の差分が閾値以上の場合に処理の開始を判定する平準化開始判定ステップと、

ディスク残量が最も小さいネットワークストレージ装置をファイルの移動元となる移動元ネットワークストレージ装置として選択し、ディスク残量が最も大きいネットワークストレージ装置をファイルの移動先となる移動先ネットワークストレージ装置として選択するネットワークストレージ装置選択ステップと、

前記移動元ネットワークストレージ装置に格納されている一つ又は複数のファイルを移動対象ファイルとして選択するファイル選択ステップと、

前記移動対象ファイルを前記移動元ネットワークストレージ装置から前記移動先ネットワークストレージ装置へ移動し、該移動対象ファイルの前記配置情報を更新するファイル移動ステップと、

前記ディスク残量の最大値と最小値の差分が閾値以上である場合に前記ディスク残量調査ステップ又は前記ネットワークストレージ装置選択ステップに戻って処理を継続することを判定し、前記ディスク残量の最大値と最小値の差分が閾値未

満となった場合に処理の終了を判定する平準化終了判定ステップから成ることを特徴とする仮想一元化ネットワークストレージシステムのディスク残量の平準化方法。

【請求項 2】

前記ディスク残量調査ステップでは、前記ディスク残量の調査結果をネットワークストレージ装置を一意に識別するネットワークストレージ識別子に対応づけてディスク残量テーブルに記録し、

前記ファイル移動ステップでは、該移動対象ファイルの移動終了後に前記移動元ネットワークストレージ装置のディスク残量に該移動対象ファイルのサイズを加算した値で前記ディスク残量テーブルを更新し、前記移動先ネットワークストレージ装置の前記ディスク残量に該移動対象ファイルのサイズを減算した値で前記ディスク残量テーブルを更新し、

前記平準化終了判定ステップでは、前記ディスク残量テーブルに記録されたディスク残量を使用することを特徴とする請求項1記載の仮想一元化ネットワークストレージシステムのディスク残量の平準化方法。

【請求項 3】

前記ネットワークストレージ装置選択ステップでは、開始からの一連の平準化処理の間に移動元ネットワークストレージ装置として選択されたネットワークストレージ装置を移動先ネットワークストレージ装置の選択対象から除外し、開始からの一連の平準化処理の間に移動先ネットワークストレージ装置として選択されたネットワークストレージ装置を移動元ネットワークストレージ装置の選択対象から除外することを特徴とする請求項 1 記載の仮想一元化ネットワークストレージシステムのディスク残量の平準化方法。

【請求項 4】

前記平準化終了判定ステップは、開始からの一連の平準化処理の間に一度以上移動元ネットワークストレージ装置として選択されたネットワークストレージ装置のディスク残量の最大値が、一度以上移動先ネットワークストレージ装置として選択されたネットワークストレージ装置のディスク残量の最小値を上回った場合に該処理の終了を判定することを特徴とする請求項 1 記載の仮想一元化ネットワー

クストレージシステムのディスク残量の平準化方法。

【請求項 5】

前記ネットワークストレージ装置選択ステップでは、ディスク残量が閾値以下のネットワークストレージ装置を前記移動先ネットワークストレージ装置の選択対象から除外することを特徴とする請求項1記載の仮想一元化ネットワークストレージシステムのディスク残量の平準化方法。

【請求項 6】

前記ネットワークストレージ装置選択ステップでは、ファイルサイズの平均値が閾値よりも大きいネットワークストレージ装置を前記移動元ネットワークストレージ装置の選択対象から除外することを特徴とする請求項1記載の仮想一元化ネットワークストレージシステムのディスク残量の平準化方法。

【請求項 7】

前記ネットワークストレージ装置の中で少なくとも一つのネットワークストレージ装置のディスク残量が閾値を下回った場合に前記平準化開始判定ステップを実施することを特徴とする請求項1記載の仮想一元化ネットワークストレージシステムのディスク残量の平準化方法。

【請求項 8】

前記平準化開始判定ステップでは、処理開始時刻を平準化開始時刻として記録し、前記平準化終了判定ステップでは、判定時刻と前記平準化開始時刻の差分が閾値よりも大きい場合には、該処理の終了を判定することを特徴とする請求項1記載の仮想一元化ネットワークストレージシステムのディスク残量の平準化方法。

【請求項 9】

前記ファイル移動ステップでは、前記移動対象ファイルの移動が終わると該移動対象ファイルのファイル識別子をファイル識別子オフセットとして記録し、前記ファイル選択ステップは、該移動元ネットワークストレージ装置に格納されているファイルの中で、前記ファイル識別子が前記ファイル識別子オフセットの次に大きいファイルを前記移動対象ファイルとして選択することを特徴とする請求項1記載の仮想一元化ネットワークストレージシステムのディスク残量の平準化方法。

【請求項 10】

前記ファイル移動ステップでは、該移動元ネットワークストレージ装置に格納されているファイルの中で、前記ファイル識別子が前記ファイル識別子オフセットの次に小さいファイルを前記移動対象ファイルとして選択することを特徴とする請求項 9 記載の仮想一元化ネットワークストレージシステムのディスク残量の平準化方法。

【請求項 11】

前記ネットワークストレージ装置選択ステップは、前記移動先ネットワークストレージ装置のディスク残量と前記移動元ネットワークストレージ装置のディスク残量の差分を記録し、前記ファイル選択ステップは、前記ディスク残量の差分に基づいて、選択する移動対象ファイルの数又は移動対象ファイルの合計サイズを決定することを特徴とする請求項 1 記載の仮想一元化ネットワークストレージシステムのディスク残量の平準化方法。

【請求項 12】

前記ファイル選択ステップでは、該移動元ネットワークストレージ装置のディスク残量と該移動対象ファイルのサイズの和をファイル移動後の移動元ネットワークストレージ装置のディスク残量の予測値とし、該移動先ネットワークストレージ装置のディスク残量と該移動対象ファイルのサイズの差をファイル移動後の移動先ネットワークストレージ装置のディスク残量の予測値とし、前記移動元ネットワークストレージ装置のディスク残量の予測値と前記移動先ネットワークストレージ装置のディスク残量の予測値の大小関係と、前記移動元ネットワークストレージ装置のディスク残量と前記移動先ネットワークストレージ装置のディスク残量の大小関係が逆転している場合は、該移動対象ファイルの選択を中止して、別のファイルを移動対象ファイルとして選択することを特徴とする請求項 1 記載の仮想一元化ネットワークストレージシステムのディスク残量の平準化方法。

【請求項 13】

前記ファイル選択ステップでは、前記移動元ネットワークストレージの前記移動対象ファイルの選択の中止頻度を記録し、前記ネットワークストレージ装置選択ステップでは、前記中止頻度が閾値以上になったネットワークストレージ装置を

前記移動元ネットワークストレージ装置の選択対象から除外することを特徴とする請求項 12 記載の仮想一元化ネットワークストレージシステムのディスク残量の平準化方法。

【請求項 14】

前記ファイル移動ステップは、前記移動先ネットワークストレージ装置に前記移動対象ファイルをコピーし、該コピー終了後に、該移動対象ファイルの最終更新時刻と該移動対象ファイルのコピーのファイル生成時刻を比較し、該移動対象ファイルの最終更新時刻よりも該移動対象ファイルのコピーのファイル生成時間の方が新しい場合は、前記移動元ストレージ装置にある該移動対象ファイルを削除して該ファイル移動ステップを終了し、該移動対象ファイルのコピーのファイル生成時刻よりも該移動対象ファイルの最終更新時刻の方が新しい場合は、前記移動先ストレージ装置にある該移動対象ファイルのコピーを削除し、再度同じステップをやり直すことを特徴とする請求項 1 記載の仮想一元化ネットワークストレージシステムのディスク残量の平準化方法。

【請求項 15】

前記ファイル移動ステップは、前記移動対象ファイルのコピーのファイル生成時刻よりも前記移動対象ファイルの最終更新時刻の方が新しい場合は、前記移動先ストレージ装置にある該移動対象ファイルのコピーを削除し、前記ファイル選択ステップに戻ることを特徴とする請求項 14 記載の仮想一元化ネットワークストレージシステムのディスク残量の平準化方法。

【請求項 16】

前記ファイル移動ステップは、前記移動対象ファイルのコピー中にも、前記移動対象ファイルのコピーのファイル生成時刻と前記移動対象ファイルの最終更新時刻を比較することを特徴とする請求項 14 記載の仮想一元化ネットワークストレージシステムのディスク残量の平準化方法。

【請求項 17】

前記ファイル選択ステップでは、前記仮想一元化ネットワークストレージシステムのファイル及びディレクトリのツリー構造を一元管理する管理ツリーに存在するファイル数を母数として発生させた乱数をファイル選択の開始番号Nとし、前

記管理ツリーを再帰探索して前記N番目に見つかったファイルをファイル選択の開始地点とし、前記開始地点から再帰探索を続行し、該移動元ネットワークストレージ装置に格納されているファイルを所定回数選択することを特徴とする請求項1記載の仮想一元化ネットワークストレージシステムのディスク残量の平準化方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、仮想一元化されたネットワークストレージシステムに関連し、特にネットワークストレージ装置の間でファイル移動を伴う、ディスク使用量の平準化方法に関する。

【0002】

【従来の技術】

従来の情報システムは情報を処理する計算機に情報を格納するストレージをローカル接続して使用していた。ところが、近年ネットワーク技術の発達に伴い、ストレージを計算機から分離してネットワークに接続し、ネットワーク経由でストレージにアクセスするネットワーク形態が普及しつつある。このようなネットワーク接続されたストレージをネットワークストレージと呼ぶ。

ネットワークストレージの代表としてSAN (Storage Area Network) ストレージとNAS(Network Attached Storage)がある。SANストレージはネットワークとして専用開発されたSANを用い、性能と信頼性は高いが非常に高価であり、主にエンタープライズ向けに使用される。一方、NASはネットワークとして最も普及しているIPネットワークを使用し、性能はSANストレージより低い、価格がSANストレージに比べて低く、また手軽に使用できる。

昨今の不況により、高価で大規模なストレージを導入するより、安価で小規模なネットワークストレージを導入し、その後の必要に応じて小規模なネットワークストレージを追加していきたいという要求がある。しかし、新しいネットワークストレージをシステムに追加する場合、新旧のネットワークストレージ間のデータ移動や、クライアント（計算機を含む）とネットワークストレージの再接続を

行なわなければならず、システム管理コストが増大することが問題となる。そこで、システム管理コストを抑制するために、複数のネットワークストレージをクライアントから仮想的に一つのネットワークストレージとして見せ、新しいネットワークストレージの追加や既存ネットワークストレージの削除を行なっても、クライアントには影響を及ぼさないネットワークストレージの仮想一元化が必要となる。

ネットワークストレージの仮想一元化技術として、いくつかの方法が開発されている。例えば、http://www.maxtor.com/products/maxattach/products/applicationSpotlights/OTG_solutionsSpotlight.htmには、プライマリストレージと呼ばれる管理サーバを兼ねたネットワークストレージがファイルの配置情報を一括管理し、ファイルの最終アクセス時刻によって格納するネットワークストレージを決定する仮想一元化方法について記載されている。新規に生成されたファイルは一定期間プライマリストレージに格納され、しばらくアクセスされない古いファイルはプライマリストレージからセカンダリストレージに移動される。クライアントからのファイルアクセスはプライマリストレージが受け付け、そのファイルがセカンダリストレージに存在する場合は、セカンダリストレージにファイルアクセスを行なうことで、クライアントからはあたかも一つのネットワークストレージのように見える。

また、DiFFS: a Scalable Distributed File System, Christos Karamanolis et al., HP Laboratories Palo Alto, HPL-2001-19, January 24, 2001には、論理ボリューム単位でファイルとディレクトリを管理する、仮想一元化方法について記載されている。この方法は、ディレクトリとファイルを管理するディレクトリエントリ内にファイル格納先の論理ボリューム識別子を記録し、このディレクトリエントリを各論理ボリュームに分散配置する。各ネットワークストレージは論理ボリューム識別子とその格納先のネットワークストレージ識別子との対応テーブルを持っており、そのテーブルとディレクトリエントリからファイル格納先のネットワークストレージを特定する。新規のネットワークストレージがシステムに追加された場合、物理ディスクの仮想化技術であるLVM(Logical Volume Manager)のミラーリング機能を使用して、既存のネットワークストレージから新

規のネットワークストレージへ論理ボリュームを移動する。

また、米国特許第 6, 0 2 9, 1 6 8 号には、一つのファイルを部分的に複数のネットワークストレージに分散配置する仮想一元化方法が記載されている。ファイルを配置するネットワークストレージの分散範囲と順序のファイル管理情報を持ち、新規のネットワークストレージが追加された場合、このファイル管理情報を更新する。更新以降の新規生成ファイルは新しい分散範囲に配置されるようになる。ただしこの方法では、既存ファイルのファイル管理情報は更新されておらず、既存のファイルまたはその一部分が新規のネットワークストレージに移動されることはない。

また、特開平 0 6 - 5 9 9 8 2 には、高速外部記憶装置のディスク残量に基づいて低速外部記憶装置にデータを退避させるかどうかを決定する、計算機の仮想記憶の制御方法について記載されている。この方法では、磁気ディスクよりも高速な高速外部記憶装置と低速だが大容量の磁気ディスクの低速外部記憶装置を備えている。高速外部記憶装置のディスク残量が閾値以下になった場合、データを外部記憶装置に退避させ、高速外部記憶装置のディスク残量が閾値以上になると低速外部記憶装置から高速外部記憶装置へデータを戻す。これにより、二つの外部記憶装置が計算機から一つの仮想記憶として見える。

【 0 0 0 3 】

【発明が解決しようとする課題】

前記 http://www.maxtor.com/products/maxattach/products/applicationSpotlights/OTG_solutionsSpotlight.htm の方法では、最終アクセス時刻によってファイルを格納するストレージを決定するため、プライマリストレージとセカンダリストレージの間で定常的なディスク使用量の不均衡が発生する。また、前記米国特許第 6, 0 2 9, 1 6 8 号では、ファイルが均等に分散配置されるため同時期にシステムに追加されたネットワークストレージ間ではディスク使用量が平準であるが、追加された時期が違うネットワークストレージ間では、ファイルの移動が行なわれないため、定常的なディスク使用量の不均衡が発生する。これらの定常的な不均衡が原因となり、あるネットワークストレージが飽和すると、他のネットワークストレージのディスク残量に余裕があるにも関わらず、ファイルへの書

込みが出来なくなることがある。

【0004】

この問題は、前記DiFFS：a Scalable Distributed File Systemの方法に、各ネットワークストレージのディスク使用率を平準化する機能を追加すれば、解決することができる。しかし、ディスク使用率が平準化されていても、各ネットワークストレージのディスク容量が不均一なシステムでは、各ネットワークストレージのディスク残量が不均一となる。ここで、ディスク残量が最小のネットワークストレージに格納されているファイルに大量の書込みが発生すると、そのストレージはすぐに飽和してしまう。この飽和によってファイルへの書込みが出来なくなることがある。

【0005】

また、このDiFFS：a Scalable Distributed File Systemの方法では、論理ボリュームの移動中にクライアントからのアクセス要求をバッファリングしている。そのため、論理ボリュームの移動中にクライアントからのアクセス要求が頻発すると、バッファフルを起こす可能性がある。このバッファフルが起ると、システムはそれ以上アクセス要求を処理することが出来なくなり、クライアントからみてアクセス処理が止まってしまう。

【0006】

前記特開平06-59982号は、外部記憶装置のディスク残量に注目している点で、ファイルへの大量書込みが発生した場合の問題を解決するヒントになる発明である。ただし、前提となるシステムが高速外部記憶と低速外部記憶の2つのみであり、複数のネットワークストレージにより構成されているストレージシステムにそのまま応用できる発明ではない。

本発明の第一の課題は、ネットワークストレージのディスク容量が不均一な仮想一元化ネットワークストレージシステムにおいて、各ネットワークストレージ間のディスク使用量の定常的な不均衡を防止することにより、クライアントが使用可能なディスク容量を最大化すると共に、ファイルへの大量書込みに対してのシステム全体のマージンを最大化する、仮想一元化されたネットワークストレージシステムのディスク残量の平準化方法を提供することにある。

【0007】

本発明の第二の課題は、ネットワークストレージ間でファイルの移動を行なう仮想一元化されたネットワークストレージシステムにおいて、ネットワークストレージ間のファイルの移動中に、クライアントからのアクセス要求を止めない、仮想一元化ネットワークストレージシステムのディスク残量の平準化方法を提供することにある。

【0008】**【課題を解決するための手段】**

本発明の第一の課題は、各ネットワークストレージのディスク残量を平準化する手段によって解決できる。この手段を具体的に言うと、各ネットワークストレージ装置のディスク残量を調査するディスク残量調査ステップと、前記ディスク残量調査ステップの結果から前記ディスク残量の最大値と最小値を求め、該最大値と最小値の差分が閾値以上の場合に処理の開始を判定する平準化開始判定ステップと、前記ディスク残量が最も小さいネットワークストレージ装置をファイルの移動元となる移動元ネットワークストレージ装置として選択し、前記ディスク残量が最も大きいネットワークストレージ装置をファイルの移動先となる移動先ネットワークストレージ装置として選択するネットワークストレージ装置選択ステップと、前記移動元ネットワークストレージ装置に格納されている一つ又は複数のファイルを移動対象ファイルとして選択するファイル選択ステップと、前記移動対象ファイルを前記移動元ネットワークストレージ装置から前記移動先ネットワークストレージ装置へ移動し、該移動対象ファイルの前記配置情報を更新するファイル移動ステップと、前記ディスク残量の最大値と最小値の差分が閾値以上である場合に前記ディスク残量調査ステップ又は前記ネットワークストレージ装置選択ステップに戻って処理を継続することを判定し、前記ディスク残量の最大値と最小値の差分が閾値未満となった場合に処理の終了を判定する平準化終了判定ステップから成る。

本発明の第二の課題は、上記のファイル移動ステップで移動中のファイルに対して、クライアントから書込み要求があった場合に、その移動を中止する手段によって解決できる。この手段を具体的に言うと、移動元ネットワークストレージ装置

から移動対象ファイルを選択後、移動先ネットワークストレージ装置に移動対象ファイルをコピーし、コピー終了後に、移動対象ファイルの最終更新時刻と移動対象ファイルのコピーのファイル生成時刻を相互に比較し、移動対象ファイルの最終更新時刻よりも該移動対象ファイルのコピーのファイル生成時間の方が新しい場合は移動元ネットワークストレージ装置にある移動対象ファイルを削除してファイルの移動ステップを終了し、移動対象ファイルのコピーのファイル生成時刻よりも移動対象ファイルの最終更新時刻の方が新しい場合は、移動先ストレージ装置にある移動対象ファイルのコピーを削除し、再度移動対象ファイルの選択をやり直す。

【0009】

【発明の実施の形態】

<実施例 1>

図 1 は本発明の第一の実施例を含む情報システムの全体構成を示す図である。クライアント 1 はネットワーク 2 を介して仮想一元化部 9 と本発明のディスク残量平準化部 10 より構成される仮想一元化装置 3 に接続されている。仮想一元化装置 3 はネットワーク 4 を介して、ネットワークストレージ 5、6、7 と接続されている。ネットワーク 4 を用意する代わりにネットワークストレージ 5、6、7 をネットワーク 2 に接続しても良い。ネットワーク 2 と 4 を独立すると、クライアントに対するアクセス処理と本発明のディスク残量平準化処理を高い性能レベルで共存させることができる。一方ネットワーク 2 にネットワークストレージ 5、6、7 を接続すると、ネットワーク 4 が不要となり、コストを抑えることができる。仮想一元化装置 3 は、仮想一元化部 9 によってクライアントからネットワークストレージ 5、6、7 を仮想的に一元化した仮想一元化ネットワークストレージシステム 8 を提供し、本発明のディスク残量平準化部 10 によって、そのネットワークストレージシステム 8 のディスク容量の有効活用が可能となる。

(ネットワークストレージ)

ネットワークストレージ 5、6、7 は公知の技術によって構成されるもので、リモート制御手段 11 とストレージ装置 12 とを備える。ストレージ装置 12 にはファイルシステムを搭載する。ネットワークストレージ 5、6、7 は専用装置で

あっても良いし、ストレージ装置を備えた汎用のサーバ、ワークステーション、あるいは、PCにリモート制御手段を搭載しても良い。リモート制御手段11は、ネットワーク上に接続されたクライアントからのマウント要求とファイルの生成、読み出し、書き込み、あるいは、ディレクトリの生成等の要求を受け、ローカルストレージ装置のファイルシステムに対してファイルの生成、読み出し、書き込み、あるいは、ディレクトリの生成等を行なう。リモート制御手段11は、サン・マイクロシステムズ社(Sun Microsystems, Inc.)によって開発されたNFS(Network File System)や、マイクロソフト社(Microsoft Corporation)によって開発されたCIFS(Common Internet File System)等のファイルアクセスプロトコルを使用して、クライアントとの通信を行なう。本実施例1ではファイルアクセスプロトコルとしてNFSを用いるものとする。NFSの要求を受けファイルアクセス処理を行なうリモート制御手段11としては、公知のmountdとnfsdを使用する。

(仮想一元化装置)

仮想一元化装置3は、公知の技術による仮想一元化部9と本発明のディスク残量平準化部10より構成される。本実施例において、仮想一元化装置3は、ファイルのデータ部分は保持しておらず、仮想一元化ストレージシステムの管理に特化した専用装置である。

【0010】

仮想一元化部9は、ファイルシステム100、要求処理手段110、管理手段120から構成される。ファイルシステム100は、仮想一元化ネットワークストレージシステム8に存在するディレクトリとファイルのツリー構造と、各ファイルの格納先ネットワークストレージを記憶している。ファイルシステム100は、マクスター社(Maxtor Corporation)によって開発されたMaxAttachのように、ファイルのデータ部分を含んでいるものでも良い。要求処理手段110は、ファイルシステム100からファイルの格納先ネットワークストレージを特定し、そのネットワークストレージのリモート制御手段11に処理要求を行なうことで、クライアント1からのファイルアクセス処理要

求を実行する。管理手段120は仮想一元化ネットワークストレージシステム8を管理する管理者からの指示を受け付け、仮想一元化装置3の設定を変更したり、ファイルシステムの構成変更等の要求に対応したりする。本発明のディスク残量平準化部10の設定情報も、管理者から管理手段120を経由して設定される。詳細は、後述の（ディスク残量平準化部の設定情報）の項で説明する。

【0011】

ファイル格納先となる各ネットワークストレージ装置5、6、7は、ファイルシステム100と同じツリー構造を持っていたとしても良いし、独自のツリー構造を持っていたとしても良い。前者の場合は、ファイルシステム100のツリー構造が破壊されても、各ネットワークストレージ5、6、7から持つツリー構造から、復元可能なシステムが実現できる。但し、本発明のディスク残量平準化処理中のファイル移動前後でファイルシステム100とネットワークストレージの間でツリー構造の一貫性を保たなければ成らない。そのため、ファイル移動中はクライアントからのディレクトリ変更要求を待たせる必要がある。後者の場合は、ツリー構造がファイルシステム100にしか存在しないため、ファイルシステム100のバックアップを取る必要がある。しかし、各ネットワークストレージが独自のツリー構造を持っているために、ファイル移動前後でファイルシステム100とネットワークストレージの一貫性を保つ必要がなく、ファイル移動中にクライアントからのディレクトリ変更要求を待たせる必要がない。システムの可用性を重視する場合は、前者の構成をとれば良いし、ディスク残量平準化処理中のアクセス処理性能を重視する場合は、後者の構成をとれば良い。

【0012】

本発明のディスク残量平準化部10は、ディスク残量監視手段150、平準化制御手段160、ファイル移動手段170から構成される。ディスク残量監視手段150と平準化制御手段160の連携に必要な情報については後述の（ディスク残量監視手段と平準化制御手段の連携に必要な情報）の項で、平準化制御手段160とファイル移動手段170の連携に必要な情報については後述の（平準化制御手段とファイル移動手段の連携に必要な情報）の項で説明する。

【0013】

ディスク残量監視手段 150 は各ネットワークストレージのディスク残量を常時監視し、必要があれば平準化制御手段 160 にディスク残量平準化処理の開始を指示する。詳細は後述の(ディスク残量監視手段)の項で説明する。平準化制御手段 160 は、ファイルの移動元及び移動先となるネットワークストレージを決定して、ファイル移動手段 170 を制御する。詳細は後述の(平準化制御手段)の項で説明する。ファイル移動手段 170 は、平準化制御手段 160 によって指定されたファイルの移動元ネットワークストレージから移動先ネットワークストレージへファイルを移動する。詳細は後述の(ファイル移動手段)の項で説明する。

(ディスク残量平準化部の設定情報)

管理手段 120 によって設定されるディスク残量平準化部の設定情報は、図 2 に示す通り、ディスク残量監視間隔 (T c h e c k) 1251、平準化実行時間 (T f l a t) 1252、不均衡ディスク残量差 (D u b a l) 1261、平準化抑止ディスク残量 (R s u p p) 1262、飽和ディスク残量 (R f u l l) 1263、最大平均ファイルサイズ (S A V R m a x) 1271、最大リトライ回数 (R T R Y m a x) 1272 からなる。T c h e c k 1251 と T f l a t 1252 はディスク残量平準化処理のスケジューリングに使用され、D u b a l 1261 と R s u p p 1262 と R f u l l 1263 はディスク残量平準化処理の開始及び終了判定に使用され、S A V R m a x 1271 と R T R Y m a x 1272 はネットワークストレージの選択に使用される。

【0014】

ディスク残量監視間隔 (T c h e c k) 1251 は、ディスク残量監視手段 150 によって参照され、ネットワークストレージのディスク残量の監視間隔を示す設定情報である。T c h e c k 1251 は、1 時間から数週間程度が妥当だと考えられるが、それ以上、またはそれ以下の間隔であってもかまわない。T c h e c k 1251 が短いほど、ディスク残量の急激な減少に対して正確に状況を把握できるようになるが、逆にディスク残量の監視処理自身が頻繁に動くため、仮想一元化装置の処理が重くなる。従って、ディスク残量の増減の度合いによって、T c h e c k 1251 を適切に設定する。平準化実行時間 (T f l a t) 12

5 2 は、平準化制御手段 1 6 0 に参照され、平準化処理の実行時間を示す設定情報である。平準化制御手段 1 6 0 は、ディスク残量平準化処理を開始してから、T f l a t 1 2 5 2 の時間が経過すると、終了条件に関係なく強制的に処理を終了する。この機能によって、例えば、クライアント 1 からのアクセス要求が少ない時間帯だけディスク残量平準化処理を実行したいという要求に応えることが可能となる。T f l a t 1 2 5 2 は数分程度から 1 日程度が妥当な値と考えられが、それ以外の値をとっても良い。

【0 0 1 5】

不均衡ディスク残量差 (D u b a l) 1 2 6 1 は、ディスク残量監視手段 1 5 0 と平準化制御手段 1 6 0 によって参照されるディスク残量の最大値と最小値の差分の閾値で、ディスク残量の不均衡がシステム内で発生しているかを判定するための設定情報である。ディスク残量の最大値と最小値の差分が D u b a l 1 2 6 1 よりも大きい場合、ディスク残量監視手段 1 5 0 によってネットワークストレージ間のディスク残量の不均衡が発生していると判定され、ディスク残量平準化処理がはじまる。一方、ディスク残量の最大値と最小値の差分が D u b a l 1 2 6 1 を下回った場合、平準化制御手段 1 6 0 にネットワークストレージ間のディスク残量の不均衡が改善されたと判定され、ディスク残量平準化処理が終了する。D u b a l 1 2 6 1 は、0 B 以上で、最もディスク容量の小さいネットワークストレージのディスク容量よりも小さい値を指定する。例えば、最もディスク容量の小さいネットワークストレージのディスク容量が 1 0 0 G B の場合、D u b a l 1 2 6 1 は 1 G B から 2 0 G B 程度までが妥当だと考えられるが、0 B から 1 0 0 G B の間であればどの値をとっても良い。平準化抑止ディスク残量 (R s u p p) 1 2 6 2 は、ディスク残量監視手段 1 5 0 と平準化制御手段 1 6 0 が参照するディスク残量の閾値で、ディスク残量平準化処理を抑止するための設定情報である。ディスク残量の最小値が D s u p p 1 2 6 2 以上であれば、ディスク残量に十分余裕があると判定され、たとえディスク残量の不均衡が起っていてもディスク残量平準化処理は抑止される。飽和ディスク残量 (R f u l l) 1 2 6 3 は、ディスク残量監視手段 1 5 0 と平準化制御手段 1 6 0 が参照するディスク残量の閾値で、ネットワークストレージが飽和している場合にディスク残量平

準化処理を抑止するための設定情報である。ディスク残量の最大値が R f u l l 1 2 6 2 を下回っていれば、ディスク残量平準化処理は抑止される。

【0016】

最大平均ファイルサイズ (S A V R m a x) 1 2 7 1 は、平準化制御手段 1 6 0 が参照し、各ネットワークストレージの平均ファイルサイズの閾値で、巨大なファイルが多数存在するネットワークストレージから他のネットワークストレージへのファイルの移動を抑止するための設定情報である。最大リトライ回数 (R T R Y m a x) 1 2 7 2 は、ファイル移動手段 1 7 0 が行なうファイル選択のリトライ回数の閾値で、S A V R m a x 1 2 7 1 と同様に巨大なファイルが多数存在するネットワークストレージから他のネットワークストレージへのファイルの移動を抑止するための設定情報である。

(ディスク残量監視手段と平準化制御手段の連携に必要な情報)

ディスク残量監視手段 1 5 0 と平準化制御手段 1 6 0 の連携に必要な情報は、図 3 に示す通り、ネットワークストレージ情報テーブル 1 5 5、平準化実行フラグ (F r u n) 1 5 6 1、システム飽和フラグ (F f u l l) 1 5 6 2 からなる。ネットワークストレージ情報テーブル 1 5 5 は、各ネットワークストレージのディスク容量の情報を格納するためのテーブルである。詳細は後述の (ネットワークストレージ情報テーブル) の項で説明する。平準化実行フラグ (F r u n) 1 5 6 1 は、ディスク残量監視手段 1 5 0 によって設定され、平準化制御手段 1 6 0 にディスク残量平準化処理の開始を指示するフラグである。システム飽和フラグは、ディスク残量監視手段 1 5 0 によって設定され、平準化制御手段 1 6 0 に全てのネットワークストレージが飽和していることを通知するフラグである。F f u l l 1 5 6 2 が 1 に設定されている場合、たとえ F r u n 1 5 6 1 が 1 に設定されていても、ディスク残量平準化処理は抑止される。

(ネットワークストレージ情報テーブル)

ネットワークストレージ情報テーブル 1 5 5 を図 4 に示す。1 5 5 1 の行は各ネットワークストレージ装置の識別子を示し、各手段がネットワーク情報テーブル 1 5 5 に読み書きを行なうためのインデックスとなる。1 5 5 2 の行は各ネットワークストレージのディスク容量 (ファイルを一つも格納しない場合の使用可能

容量)を示す。1553の行は各ネットワークストレージの現在のディスク残量を示す。この行はディスク残量監視手段150によって定期的に設定され、平準化制御手段160がファイルの移動方向及びデータ移動量の決定、処理の終了判定を行なうために使用される。例えば図2でディスク容量の単位をGBと定めると、ネットワークストレージ1、2、3、4のディスク容量はそれぞれ120GB、100GB、200GB、90GBであり、ディスク残量はそれぞれ30GB、20GB、80GB、90GBである。1554の行は各ネットワークストレージの平均ファイルサイズを示し、ファイルの移動元となるネットワークストレージを選択する際に使用される。例えば図2で平均ファイルサイズの単位をkBと定めると、ネットワークストレージ1、2、3、4の平均ファイルサイズは、それぞれ155kB、122kB、3500kB、800kBである。仮に、最大平均ファイルサイズ(SAVRmax)1271を1000kBとすると、No.3のネットワークストレージは平準化処理中にファイルの移動元として選択されない。

(平準化制御手段とファイル移動手段の連携に必要な情報)

平準化制御手段160とファイル移動手段170の連携に必要な情報は、図5に示すとおり、ネットワークストレージ属性テーブル165、移動元ネットワークストレージ番号(Ns)1661、移動先ネットワークストレージ番号(Nd)1662、最大移動データ量(Qmax)1663、最小移動データ量(Qmin)1664、平準化終了予定時刻(Tend)1665からなる。ネットワークストレージ属性テーブル165は、各ネットワークストレージがファイルの移動先又はファイルの移動元として選択可能かどうかを示す属性情報を記録するテーブルで、ネットワークストレージ間のファイルの移動方向を固定したり、巨大なファイルを格納しているネットワークストレージからのファイルの移動を抑止するために使用される。詳細は後述の(ネットワークストレージ属性テーブル)の項で説明する。移動元ネットワークストレージ番号(Ns)1661は、ファイルの移動元となるネットワークストレージの識別番号である。移動先ネットワークストレージ番号(Nd)1662はファイルの移動先となるネットワークストレージの識別番号である。最大データ移動量(Qmax)1663は、ファイ

ル移動手段 170 が一度に移動するファイルサイズの合計の最大値である。最小データ移動量 (Q_{min}) 1664 は、ファイル移動手段 170 が一度に移動するファイルサイズの合計の最小値である。平準化終了予定時刻 (T_{end}) 1665 は、平準化処理が終了する予定時刻で、平準化処理の開始時刻に平準化処理実行時間 (T_{flat}) 1252 を加算した値である。ファイル移動手段 170 は、ファイル移動中であっても T_{end} 1665 を経過すると処理を終了する。

(ネットワークストレージ属性テーブル)

ネットワークストレージ属性テーブルを図 6 に示す。1651 の行は、各ネットワークストレージ装置の識別子を示し、各手段がネットワーク属性テーブル 165 に読み書きを行なうためのインデックスとなる。1652 の行は、各ネットワークストレージが移動元ネットワークストレージとして選択可能かどうかを示す。1652 の行に “NULL” が記述されていれば、移動元ネットワークストレージとして選択可能で、“Don't Select” が記述されていれば移動元ネットワークストレージとして選択されない。図 6 の例では、No. 3 と No. 4 が移動元ネットワークストレージとして選択されないネットワークストレージである。1652 の行は、各ネットワークストレージが移動先ネットワークストレージとして選択可能かどうかを示す。1653 の行に “NULL” が記述されていれば、移動元ネットワークストレージとして選択可能で、“Don't Select” が記述されていれば移動先ネットワークストレージとして選択されない。図 6 の例では、No. 2 と No. 4 のネットワークストレージが移動先ネットワークストレージとして選択されない。特に No. 4 のネットワークストレージは移動元ネットワークストレージとしても移動先ネットワークストレージとしても選択されない。

(ディスク残量監視手段)

ディスク残量監視手段 150 は定期的に各ネットワークストレージのディスク残量を計測し、平準化制御手段 160 に平準化処理の開始又は抑止を指示する。図 7 にディスク残量監視手段の処理フローを示す。

【0017】

まず最初に、1501 でディスク残量監視手段 150 が起動され、1502 で

平準化実行フラグ (F r u n) 1561 とシステム飽和フラグ (F f u l l) 1562 を 0 に初期化する。次に 1503 で各ネットワークストレージの平均ファイルサイズを計測し、計測結果をネットワークストレージ情報テーブル 155 の平均ファイルサイズの行 1554 に書込む。この情報は後述の平準化制御手段 160 がネットワークストレージ属性テーブル 165 を初期化するために使用する。次に 1504 で各ネットワークストレージのディスク残量を計測し、計測結果をネットワークストレージ情報テーブル 155 のディスク残量の行 1553 に書込む。ディスク残量の計測は、各ネットワークストレージに NFS プロトコルの S T A T F S プロシージャを発行しても良いし、その他ネットワーク経由で各ネットワークストレージのディスク残量を取得できればどのような方法を用いても良い。

【0018】

次に、1505 では、1504 の計測結果から、ネットワークストレージのディスク残量の最大値 (R m a x) とディスク残量の最小値 (R m i n) を求める。次の 1506 で、ディスク残量の最大値 (R m a x) が飽和ディスク残量 (R f u l l) 1263 未満の場合 (ファイルシステムが飽和している場合)、1508 に移ってシステム飽和フラグ (F f u l l) 1562 を 1 に設定して、1509 ~ 1511 を行なわずに 1512 に移る。ディスク残量の最大値 (R m a x) が飽和ディスク残量 (R f u l l) 1263 以上である場合 (ファイルシステムが飽和していない場合)、1507 に移ってシステム飽和フラグ (F f u l l) 1562 を 0 に 0 に設定して、次の 1509 に移る。1509 では、ディスク残量の最小値 (R m i n) が平準化抑止ディスク残量 (R s u p p) 1262 以上の場合、1510、1511 を行なわずに 1512 に移る。ディスク残量の最小値 (R m i n) が平準化抑止ディスク残量 (R s u p p) 1262 未満の場合、次の 1510 に移る。1510 で、ディスク残量の最大値 (R m a x) と最小値 (R m i n) の差分が不均衡ディスク残量差 (D u b a l) 1261 未満の場合 (ディスク残量の不均衡が起っていない場合)、1511 を行なわずに 1512 に移る。ディスク残量の最大値 (R m a x) と最小値 (R m i n) の差分が不均衡ディスク残量差 (D u b a l) 1261 以上の場合 (ディスク残量の不均衡

が起っている場合)、1511で平準化実行フラグ(F r u n) 1562を1に設定して、平準化制御手段160に平準化処理の開始を指示する。1512では1503を行なった時刻からディスク残量監視間隔(T c h e c k) 1251を加算した時刻までスリープし、1503の処理へ戻る。

(平準化制御手段)

平準化制御手段160は、ディスク残量監視手段150が設定する平準化実行フラグ(F r u n) 1561とシステム飽和フラグ(F f u l l) 1562を常時監視しており、それらのフラグが適切に設定されるとディスク残量平準化処理を開始する。処理開始後、平準化制御手段160は、ファイルの移動方向及び移動量を決定や、ファイル移動手段170の起動、終了条件の判定を行なう。図8に平準化制御手段160の処理フローを示す。

【0019】

1601で平準化制御手段160が起動されると、1602でF r u n 1561とF f u l l 1562の監視を始める。ディスク残量に不均衡が発生していない場合はF u b a l 1561が0に設定されているため、1602のループを繰返し、ディスク残量平準化処理を開始しない。また、システムが飽和している場合は、F f u l l 1562が1に設定されているため、同様に1602のループを繰返す。システムが飽和していない状態で(F f u l l 1562==0)、ディスク残量に不均衡が生じた場合(F u b a l 1561==1)、1603に移る。

【0020】

1603では、平準化開始時刻を計測し、その計測結果に平準化処理実行時間(T f l a t) 1252を加算して平準化終了予定時刻(T e n d) 1665を算出する。次に1604でネットワークストレージ属性テーブル165の初期化を行なう。1604の詳細フローは、後述の(ネットワークストレージ属性テーブルの初期化)の項で説明する。

【0021】

次に1605で移動元ネットワークストレージ(N s) 1661と移動先ネットワークストレージ(N d) 1662を選択する。1605の詳細フ

ローは、後述の（ネットワークストレージの選択）の項で説明する。

【0022】

1606では、移動元ネットワークストレージ（Ns）1661と移動先ネットワークストレージ（Nd）1662が正常に選択されたかどうかを判定する。Ns1661とNd1662が選択されなかった場合、1607～1611を行わず、1612に移って平準化処理実行フラグ（Frun）を0に設定して、ディスク残量平準化処理を終了する。Ns1661とNd1662が選択された場合、1607に移る。

【0023】

1607では、最大データ移動量（Qmax）1663と最小データ移動量（Qmin）1664を求める。移動元ネットワークストレージ（Ns）1661のディスク残量をRs、移動先ネットワークストレージ（Nd）1662のディスク残量をRdとすると、Qmax1663は $(Rd - Rs) / 2$ に、Qmin1664は $Qmax1663 - D ub a l l 2 6 1 / 2$ となる。

【0024】

次に1608では、ファイル移動手段170を起動し、移動元ネットワークストレージ（Ns）1661から移動先ネットワークストレージ（Nd）1662へ、合計サイズがQmin1664を超えるまでファイルを移動する。このとき、ファイル移動手段170は、合計サイズがQmaxを超えないようにファイルの移動を行なう。

【0025】

次に1609では、各ネットワークストレージのディスク残量を計測し、計測結果をディスク残量の行1553に上書きする。次に1610では、1609で更新したディスク残量の行1553を参照してネットワークストレージ属性テーブル165の更新を行なう。1610の詳細フローについては後述の（ネットワークストレージ属性テーブルの更新）の項で説明する。

【0026】

次に1611で平準化処理を継続するか終了するかを判定する。ディスク残量の最大値（Rmax）とディスク残量の最小値（Rmin）の差分が不均衡ディ

スク残量差 (D u b a l) 1 2 6 1 未満となった (ディスク残量の不均衡が解消された) 場合、ディスク残量平準化処理を終了し、1 6 1 2 に移り平準化処理実行フラグ (F r u n) を 0 に設定し、1 6 0 2 に戻る。その他、時刻が平準化終了予定時刻 (T e n d) 1 6 6 5 を経過した場合、ディスク残量の最小値 (R m i n) が平準化抑止ディスク残量 (R s u p p) 1 2 6 2 以上となった (ディスク残量に余裕が出来たために平準化処理を行なう必要がなくなった) 場合、ディスク残量の最大値 (R m a x) が飽和ディスク残量 (R f u l l) 1 2 6 3 未満になった (全てのネットワークストレージが飽和してしまい平準化処理が続行不能となった) 場合、平準化処理を終了して 1 6 1 2 に移る。以上の条件が一つも成立しない場合、1 6 0 5 に戻り平準化処理を続行する。(ネットワークストレージ属性テーブルの初期化)

平準化制御手段 1 6 0 の 1 6 0 4 で行なうネットワークストレージ属性テーブルの初期化の詳細フローを図 9 に示す。1 6 0 4 1 で初期化開始後、1 6 0 4 2 でネットワークストレージ属性テーブルの 1 6 5 2 と 1 6 5 3 の行に “N U L L” を書込む。次に 1 6 0 4 3 で、ネットワークストレージ情報テーブル 1 5 5 の平均ファイルサイズの行 1 5 5 4 を参照し、平均ファイルサイズが最大平均ファイルサイズ (S A V R m a x) 1 2 7 1 以上であるネットワークストレージについては、1 6 5 2 の行の対応する部分に “D o n ’ t S e l e c t” を記入する。これによって平均ファイルサイズの大きいネットワークストレージは移動元ネットワークストレージとして選択されなくなる。次に 1 6 0 4 4 で、ネットワークストレージ情報テーブル 1 5 5 のディスク残量の行 1 5 5 3 を参照し、ディスク残量が飽和ディスク残量 (R f u l l) 1 2 6 3 未満のネットワークストレージについては、1 6 5 3 の行の対応する部分に “D o n ’ t S e l e c t” を記入する。これにより飽和したネットワークストレージが移動先ネットワークストレージとして選択されなくなる。

(ネットワークストレージの選択)

平準化制御手段 1 6 0 の 1 6 0 5 で行なうネットワークストレージの選択の詳細フローを図 10 に示す。1 6 0 5 1 で選択開始後、1 6 0 5 2 では、ネットワークストレージ属性テーブルの行 1 6 5 2 に “D o n ’ t S e l e c t” と記入

されていないネットワークストレージの中で、ディスク残量が最小のネットワークストレージの識別番号を移動元ネットワークストレージ番号 (Ns) 1661 に代入する。次に16053では、ネットワークストレージ属性テーブルの行1653のNs1661に対応する欄に“Don't Select”を書き込む。

【0027】

次に16054では、ネットワークストレージ属性テーブル165の行1653に“Don't Select”と記入されていないネットワークストレージの中で、ディスク残量が最大のネットワークストレージの識別番号を移動先ネットワークストレージ番号 (Nd) 1662に代入する。次に16055では、ネットワークストレージ属性テーブル165の行1652のNd1662に対応する欄に“Don't Select”を書き込み、16056で処理を終了する。

(ネットワークストレージ属性テーブルの更新)

平準化制御手段160の1610で行なうネットワークストレージ属性テーブルの更新の詳細フローを図11に示す。16101にて更新開始後、16102では、ネットワークストレージ情報テーブル155のディスク残量の行1553を参照し、ディスク残量が飽和ディスク残量 (Rfull) 1263未満のネットワークストレージについては、1653の行の対応する部分に“Don't Select”を記入する。これにより平準化処理中に飽和したネットワークストレージが移動先ネットワークストレージとして選択されなくなる。

(ファイル移動手段)

ファイル移動手段170は、平準化制御手段160から起動され、移動元ネットワークストレージ (Ns) 1661から移動先ネットワークストレージ (Nd) 1662へ、合計サイズが最小データ移動量 (Qmin) 1664以上になるまでファイルを移動する。このファイル移動手段170は2つの大きな特徴をもつ。一つ目の特徴は、移動元ネットワークストレージのディスク残量と移動先ネットワークストレージのディスク残量が逆転しないようにファイルの選択を行なうことである。(移動するファイルサイズの合計が最大データ移動量Qmax16

63未満になる様にファイルの選択を行なう。) この特徴によって、ディスク残量の振動を抑制し、無駄なファイルの移動を防ぐ。2つ目の特徴は、移動中のファイルに対してクライアントから書込み要求があった場合に、クライアントの書込み要求を優先的に処理し、ファイルの移動を破棄して、再度ファイルの移動を行なうことである。図12にファイル移動手段170の処理フローを示す。

【0028】

1701でファイル移動手段170が起動された後、1702でデータ移動量計算用の内部カウンタQを0に初期化する。次に1703では、ファイルシステム100に存在するファイル総数を母数とする乱数をオフセットNとし、N-1番目のファイルが見つかるまでファイルシステム100を再帰探索し、ファイル選択のスタート地点を決定する。

【0029】

次に1703で決定したファイル選択のスタート地点からファイルシステム100の再帰探索を継続し、Nsに格納されていて、且つサイズが最大データ移動量(Qmax)1663と内部カウンタQの差分未満のファイルを移動対象ファイルとして選択する。1704の詳細については、(移動対象ファイルの選択)の項で説明する。

【0030】

次に1705で移動対象ファイルを正常に選択できたかどうかを判定する。移動対象ファイルを選択できなかった場合、1707に移り、ネットワークストレージ属性テーブル165の行1652のNs1661に対応する欄に“Don't Select”を記入し、処理を終了する。1652に“Don't Select”と記入されたネットワークストレージは、ネットワークストレージの選択1605において移動元ネットワークストレージとして選択されない。移動対象ファイルが正常に選択された場合、次の1706に移る。1706では移動元ネットワークストレージ(Ns)1661から移動先ネットワークストレージ(Nd)1662へ、移動対象ファイルを移動する。ファイルの移動の詳細フローについては、後述の(移動対象ファイルの移動)の項で説明する。

【0031】

次に 1708 で移動対象ファイルが正常に移動されたか、それともクライアントからの書込みがおこり移動対象ファイルの移動が途中で終了したかを判定する。移動対象ファイルの移動が途中で終了した場合、1709 は行なわず 1710 に移る。移動対象ファイルを正常に移動できた場合、1709 に移り、データ移動量の内部カウンタ Q に移動対象ファイルのサイズを加算して更新する。次の 1710 で、データ移動量の内部カウンタ Q が最小データ移動量 (Q_{min}) 1664 以上である場合、又は、平準化終了予定時刻 (T_{end}) 1665 を経過している場合、ファイルの移動を終了する。それ以外の場合は、1704 に戻りファイルの移動を継続する。

(移動対象ファイルの選択)

ファイル移動手段 170 が 1704 で行なう移動対象ファイルの選択の詳細フローを図 13 に示す。

【0032】

17041 にて移動対象ファイルの選択開始後、17042 でファイル選択のリトライ回数をカウントするリトライカウンタ RC を 0 に初期化する。次に 17043 では、再帰探索にてファイルを探査する。17044 では、探索したファイルの格納先が移動元ネットワークストレージ (N_s) 1661 と一致しているかを調べ、一致していれば次の処理に移る。一致していなければ 17043 に戻り、格納先が N_s 1661 と一致するまで同じ処理を繰返す。次に 17045 で、探索したファイルの移動後に、ディスク残量の逆点が起らないかどうか判定する。探索したファイルサイズが最大データ移動量 (Q_{max}) 1663 とデータ移動量 Q の差分未満である場合、ディスク残量の逆転は起らないので、17046 で移動対象ファイルとして選択し処理を終了する。一方、探索したファイルのサイズが最大データ移動量 (Q_{max}) 1663 とデータ移動量 Q の差分以上である場合、そのファイルを移動対象ファイルとして選択しない。この場合、17047 に移りリトライカウンタ RC をインクリメントする。次に、17048 で RC が最大リトライ回数 (R_{TRYmax}) 1272 未満であれば、17043 に戻りファイルの再度ファイルの選択を行なう。17048 で RC が最大リトライ回数 (R_{TRYmax}) 1272 以上であれば、移動対象ファイルの選択を行

なわずに処理を終了する。

【0033】

(移動対象ファイルの移動)

ファイル移動手段170が1706で行なう移動対象ファイルの移動の詳細フローを図14に示す。

17061でファイルの移動を開始すると、17062に移り、移動元ネットワークストレージ(Ns)1661から移動先ネットワークストレージ(Nd)1662へ、移動対象ファイルをコピーする。次に17063で、移動対象ファイルのコピー中にクライアントからの書込みが行なわれたかどうかを判定する。コピーファイルの生成時刻が移動対象ファイルの最終更新時刻よりも新しい場合、クライアント1からの書き込みは行なわれていないので、17064に移り、ファイルシステム100の移動対象ファイルの格納先を移動元ネットワークストレージ(Nd)1661から移動先ネットワークストレージ(Ns)1662に更新する。次に17066で移動元ネットワークストレージ(Nd)1661から移動対象ファイルを削除し、ファイルの移動を終了する。

一方、17063において移動対象ファイルの最終更新時刻の方がコピーファイルの生成時刻よりも新しい場合、クライアント1からの書込みがあったということなので、ファイルの移動を行なわない。この場合、17065で移動先ネットワークストレージ(Nd)1662からコピーファイルを削除し移動対象ファイルの移動を終了する。

<実施例2>

図15は本発明の第2の実施例を含む情報システムの全体構成を示す図である。本実施例は図1の実施例1の変形例であるから、実施例1との相違部分についてのみ説明する。

仮想一元化装置13は図1の仮想一元化装置3と同等の機能を持つ。仮想一元化装置13は、ネットワークストレージ5も同一装置内に含んでいる。ここでネットワークストレージ5は仮想一元化装置13の内部のネットワークに接続されており、第1の実施例と同様に制御することが出来る。そのため、ネットワーク4に他のネットワークストレージが接続されていなく、仮想一元化装置13のみの

構成でもネットワークストレージとして機能することが可能となる。

＜実施例 3＞

図 16 は本発明の第 3 の実施例を含む情報システムの全体構成を示す図である。本実施例は図 1 の実施例 1 の変形例であるから、実施例 1 との相違部分についてのみ説明する。

仮想一元化装置 14 は図 1 の仮想一元化装置 3 と、平準化部 15 は図 1 のディスク使用量の平準化部 10 と同等の機能を持つ。但し、平準化部 15 からはファイル移動手段 170 が除いてあり、その代わり各ネットワークストレージ 16、17、18 にファイル移動手段 170 が含まれている。平準化制御手段 160 は、ネットワーク 4 を経由して、各ネットワークストレージ 16、17、18 内にあるファイル移動手段 170 を制御する。本実施例によって、仮想一元化装置 14 が行なう処理が低減されるため、結果として図 1 の第一の実施例よりも平準化処理性能と通常のアクセス処理を高い性能レベルで実現することが可能となる。

【0034】

【発明の効果】

本発明の仮想一元化ネットワークストレージシステムのディスク残量の平準化方法は、各ネットワークストレージのディスク残量を調査し、各ネットワークストレージ間でファイルを移動し、ディスク残量を平準化する。これによって、仮想一元化ネットワークストレージで起こる定常的なディスク使用量の不均衡が解消される。ネットワークストレージのディスク容量が不均一な場合であっても、各ネットワークストレージのディスク残量が平準なため、ファイルへの大量書込みに対して、局所的なネットワークストレージの飽和が起こりにくいシステムが実現される。結果として、仮想一元化ネットワークストレージシステムのディスク容量を有効利用できるようになる。

上記の平準化処理では、移動元ネットワークストレージから移動先ネットワークストレージへ移動対象ファイルをコピーし、その後移動元ネットワークストレージの移動対象ファイルを削除することによってファイル移動が実現されている。ファイル移動では、移動対象ファイルのコピー後に移動先ネットワークストレージのコピーファイルの生成時間と、移動元ネットワークストレージのコピー元フ

ファイルの最終更新時刻を比較する。この比較により、コピー中にクライアントからの書込みが行なわれたどうかを判定する。書込みが行なわれた場合は、その移動を中止し、別のファイルの移動を行なう。このように、クライアントからの書込み要求を待たせるのではなく、ファイル移動を中止する。そのため、平準化処理中においてもクライアントに対するファイル共有サービスを止めることがない。

【図面の簡単な説明】

【図 1】

本発明の第 1 実施例を含む仮想一元化ネットワークストレージシステムの全体構成を示す図である。

【図 2】

管理手段によって設定されるディスク残量平準化部の設定情報を示す図である。

【図 3】

ディスク残量監視手段と平準化制御手段の連携に必要な情報を示す図である。

【図 4】

ディスク残量監視手段と平準化制御手段の連携に使用されるネットワークストレージ情報テーブルを示す図である。

【図 5】

平準化制御手段とファイル移動手段の連携に必要な情報を示す図である。

【図 6】

平準化制御手段とファイル移動手段の連携に使用されるネットワークストレージ属性テーブルを示す図である。

【図 7】

ディスク残量監視手段が行なう処理フローを示す図である。

【図 8】

平準化制御手段が行なう処理フローを示す図である。

【図 9】

平準化制御手段が行なうネットワークストレージ属性テーブルの初期化の詳細フローを示す図である。

【図 10】

平準化制御手段が行なうネットワークストレージの選択の詳細フローを示す図である。

【図 11】

平準化制御手段が行なうネットワークストレージ属性テーブルの更新の詳細フローを示す図である。

【図 12】

ファイル移動手段が行なう処理フローを示す図である。

【図 13】

ファイル移動手段が行なう移動対象ファイルの選択の詳細フローを示す図である。

【図 14】

ファイル移動手段が行なう移動対象ファイルの移動の詳細フローを示す図である。

【図 15】

本発明の第2実施例を含む仮想一元化ネットワークストレージシステムの全体構成を示す図である。

【図 16】

本発明の第3実施例を含む仮想一元化ネットワークストレージシステムの全体構成を示す図である。

【符号の説明】

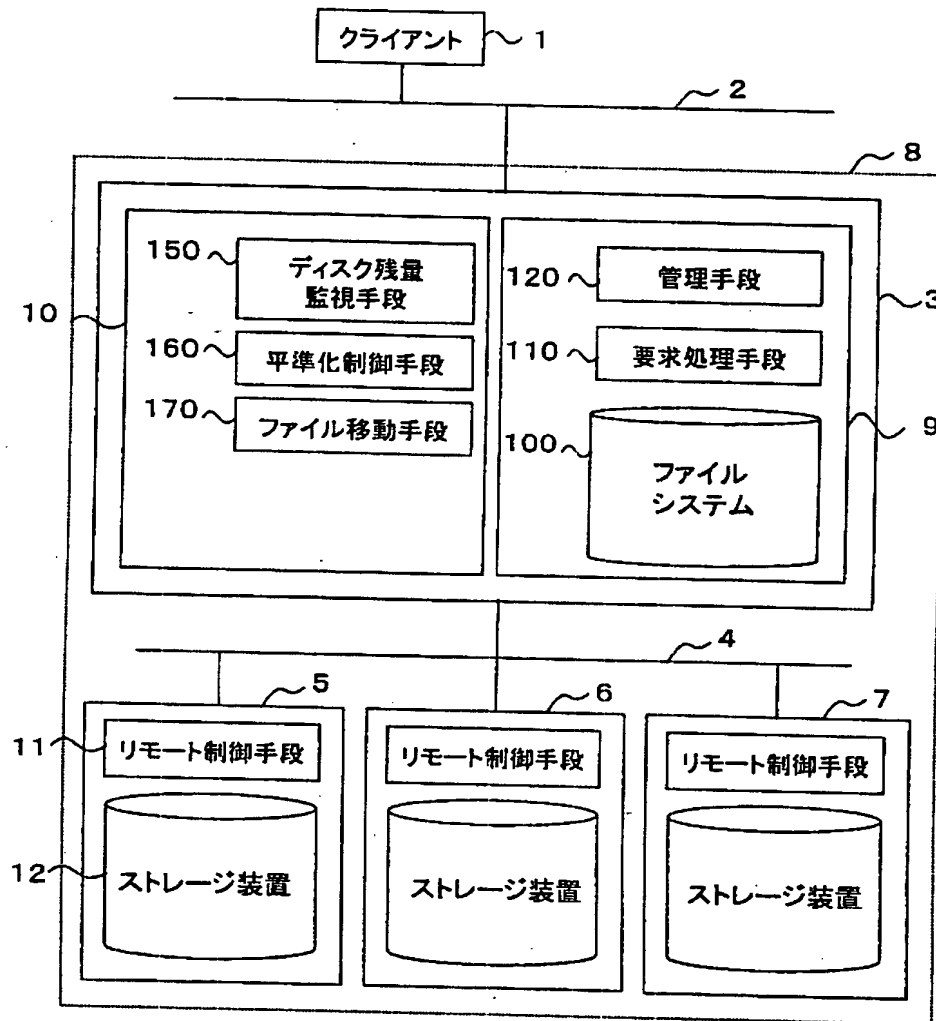
- 1 クライアント
- 2, 4 ネットワーク
- 3、13、14 仮想一元化装置
- 5、6、7、16、17、18 ネットワークストレージ装置
- 8 仮想一元化ネットワークストレージシステム
- 9 仮想一元化部
- 10 ディスク残量平準化部
- 11 リモート制御手段

- 1 2 ストレージ装置
- 1 0 0 ファイルシステム
- 1 1 0 要求処理手段
- 1 2 0 管理手段
- 1 5 0 ディスク残量監視手段
- 1 5 5 ネットワークストレージ情報テーブル
- 1 6 0 平準化制御手段
- 1 6 5 ネットワークストレージ属性テーブル
- 1 7 0 ファイル移動手段
- 1 2 5 1 ディスク残量監視間隔
- 1 2 5 2 平準化実行時間
- 1 2 6 1 不均衡ディスク残量差
- 1 2 6 2 平準化抑止ディスク残量
- 1 2 6 3 飽和ディスク残量
- 1 2 7 1 最大平均ファイルサイズ
- 1 2 7 2 最大リトライ回数
- 1 5 6 1 平準化実行フラグ
- 1 5 6 2 システム飽和フラグ
- 1 6 6 1 移動元ネットワークストレージ番号
- 1 6 6 2 移動先ネットワークストレージ番号
- 1 6 6 3 最大データ移動量
- 1 6 6 4 最小データ移動量
- 1 6 6 5 平準化終了予定時刻。

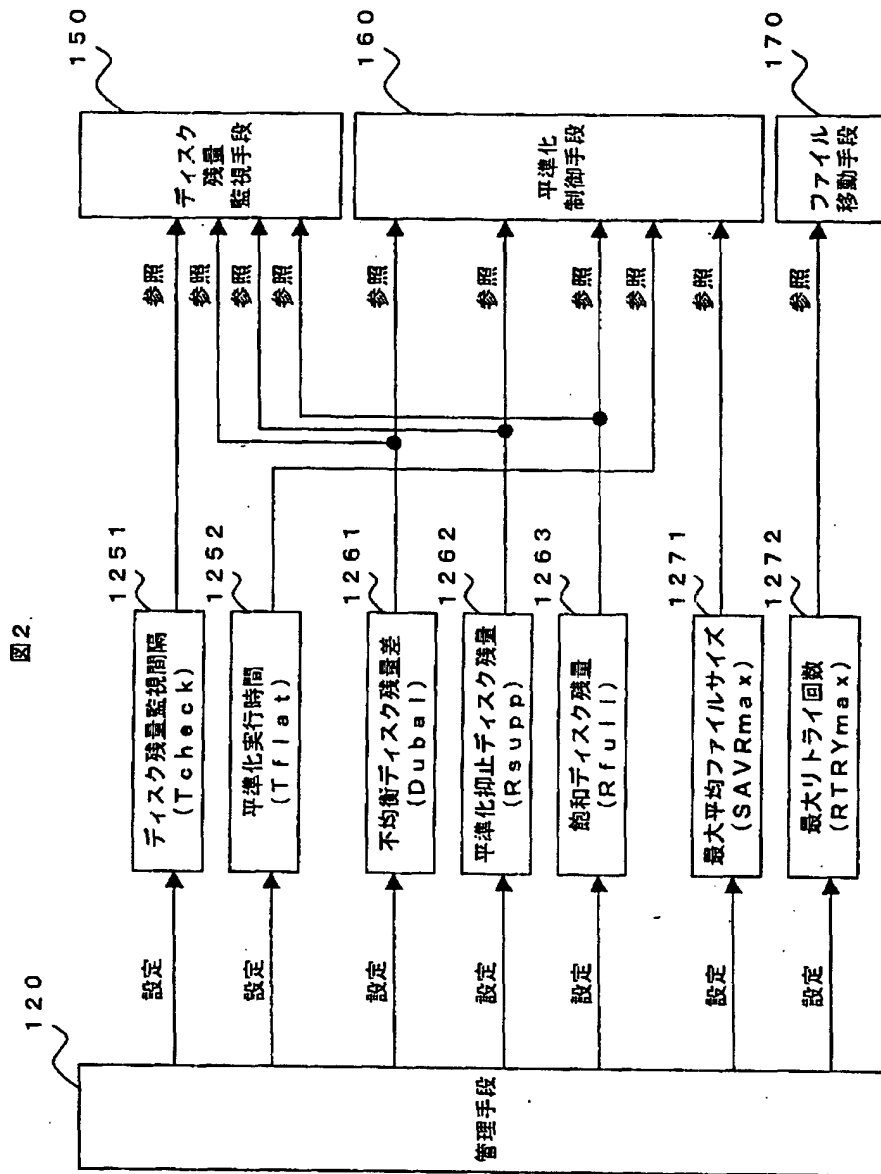
【書類名】 図面

【図 1】

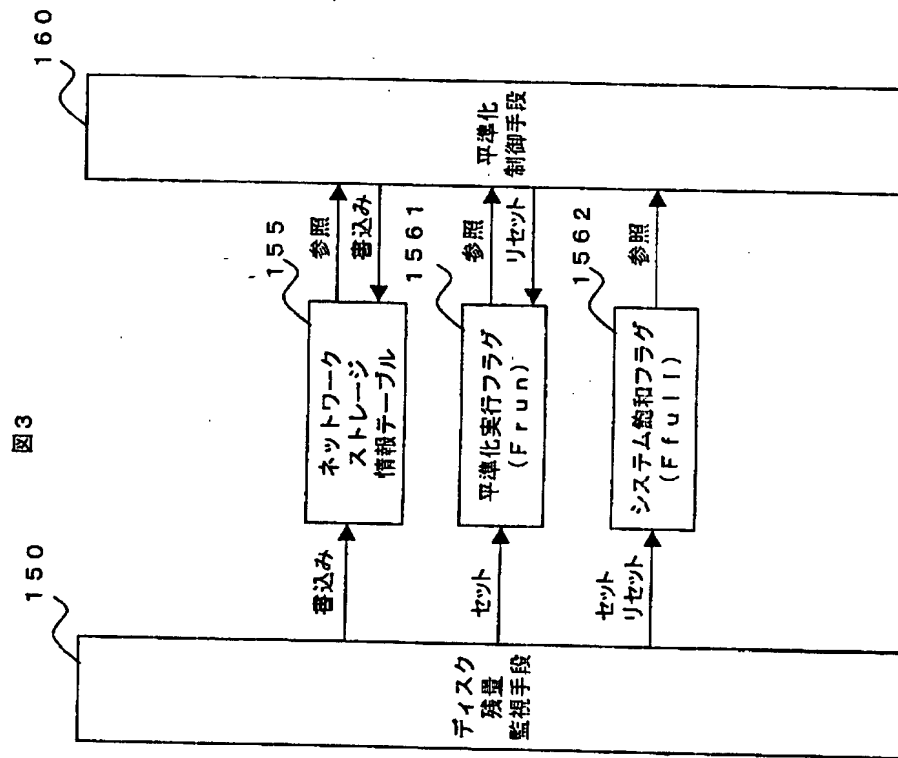
図1



【図 2】



【図 3】



【図 4】

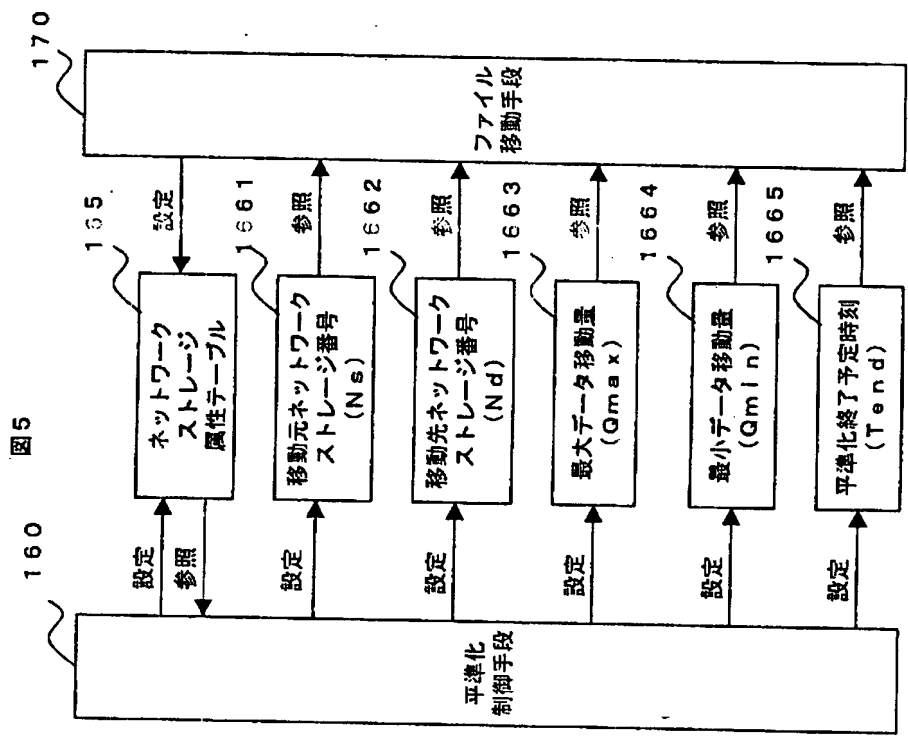
図 4

ネットワークストレージ情報テーブル

155

1551	1	2	3	4	5	...
1552	120	100	200	180	-	...
1553	30	20	80	90	-	...
1554	155	122	3500	800	-	...

【図 5】



【図 6】

図 6

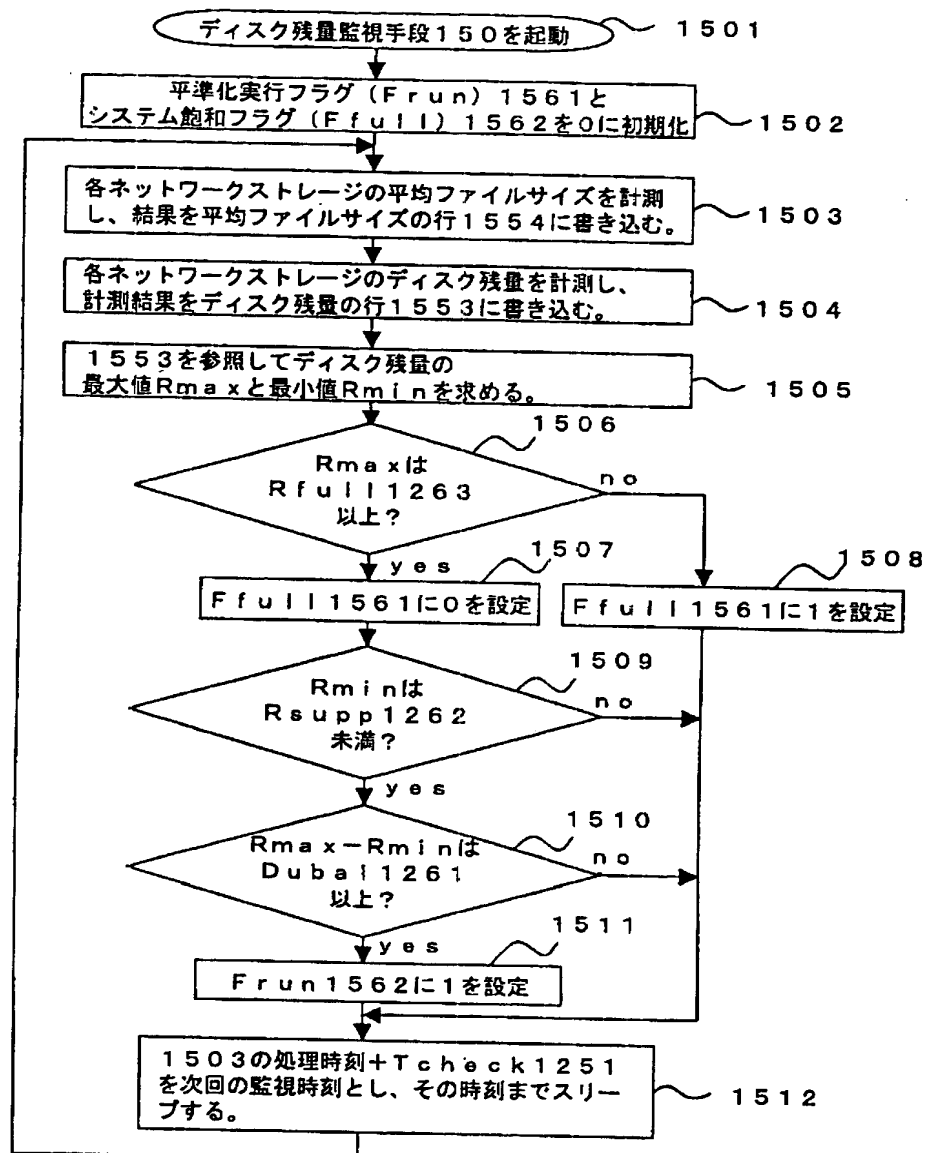
ネットワークストレージ属性テーブル

165

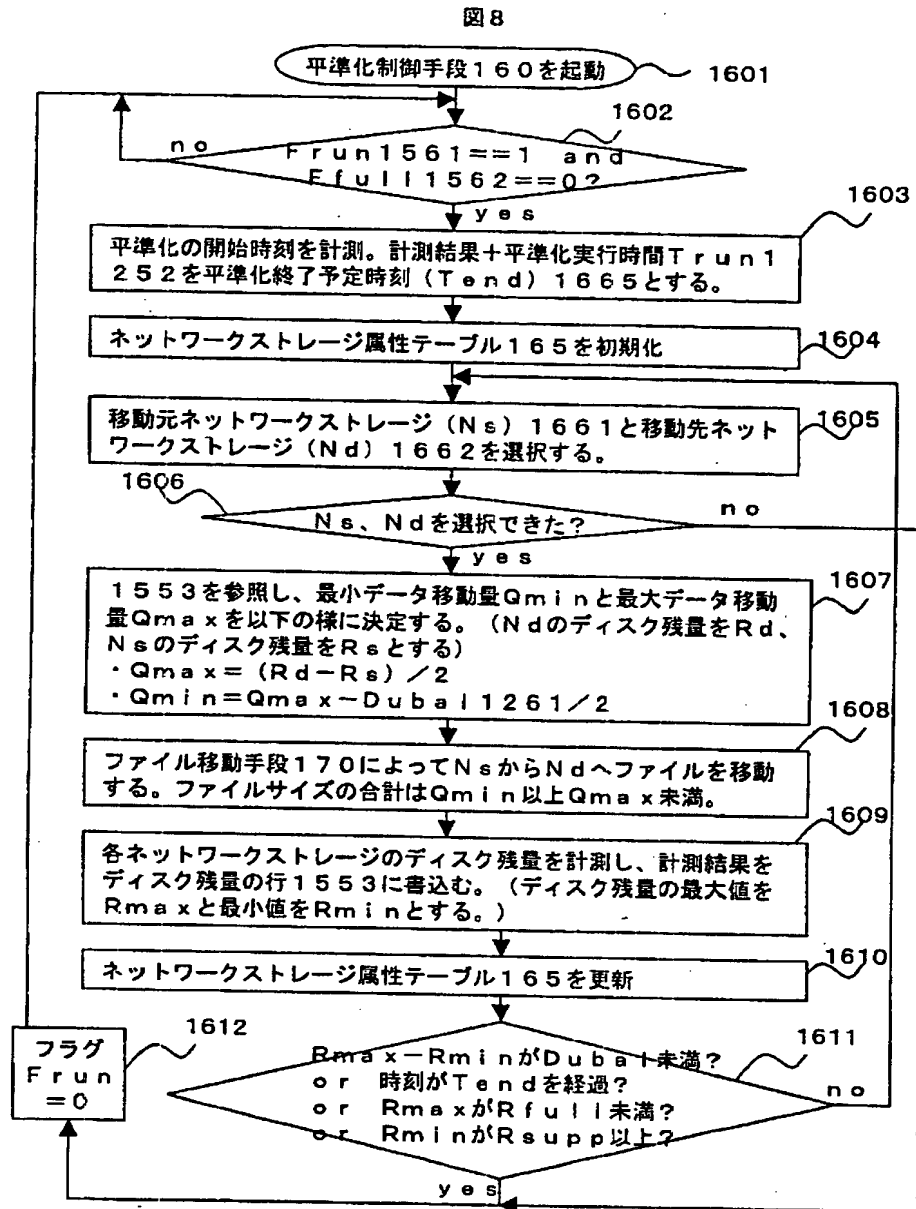
1651	1	2	3	4	5	...
1652	NULL	NULL	Don't Select	Don't Select	-	...
1653	NULL	Don't Select	NULL	Don't Select	-	...

【図 7】

図 7

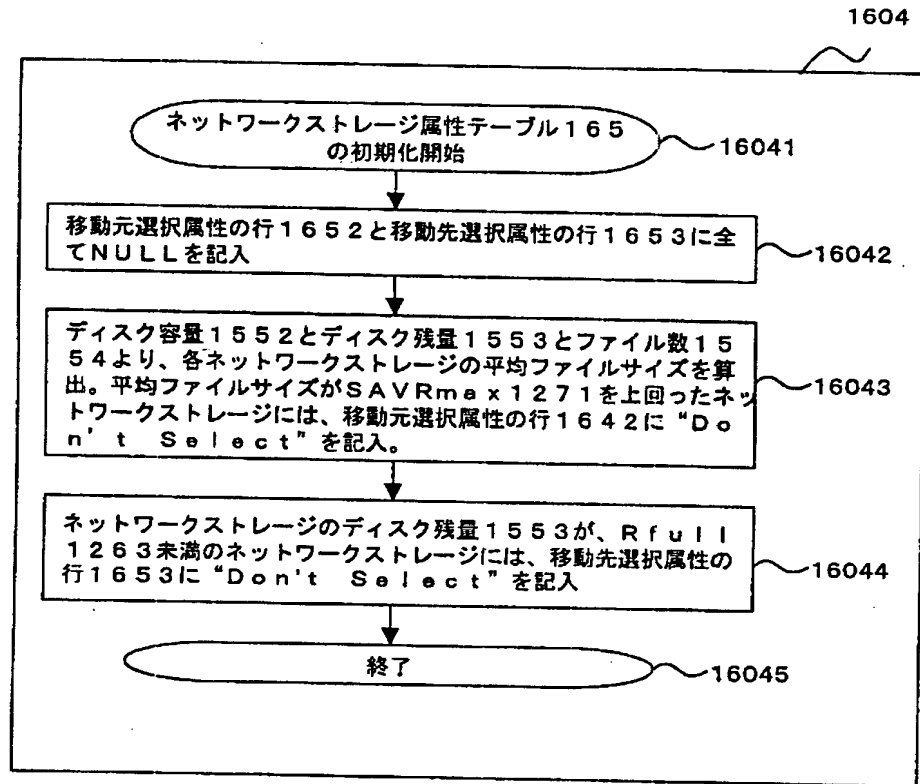


【図 8】



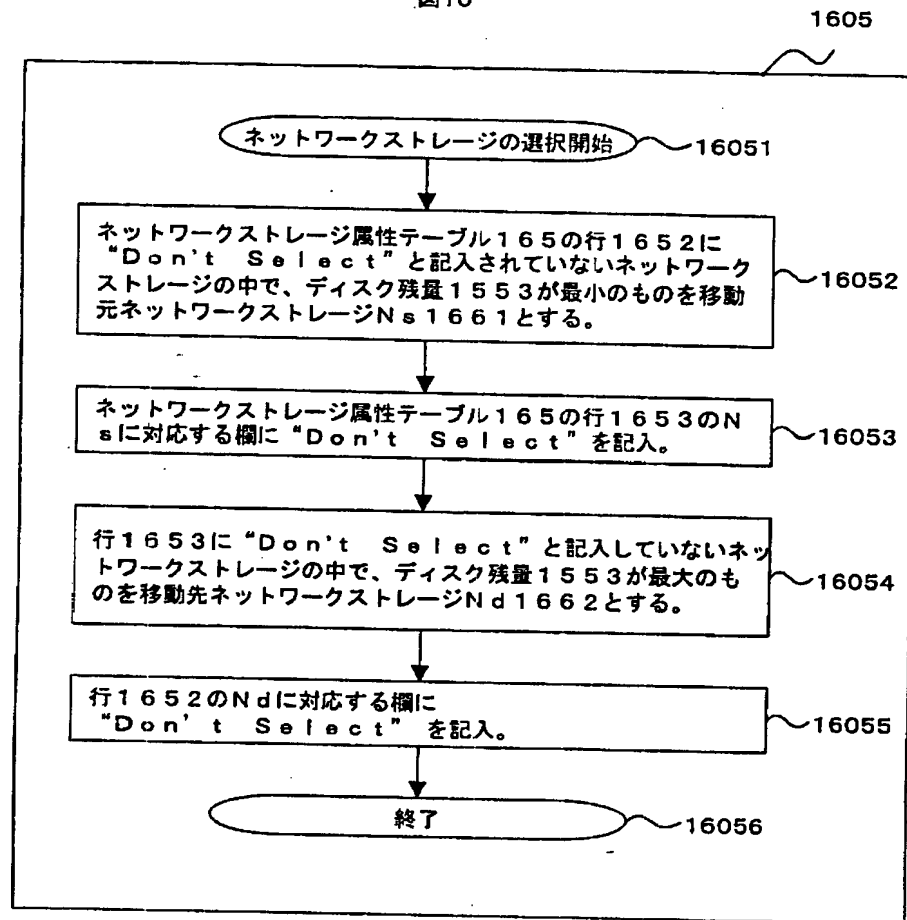
【図 9】

図9



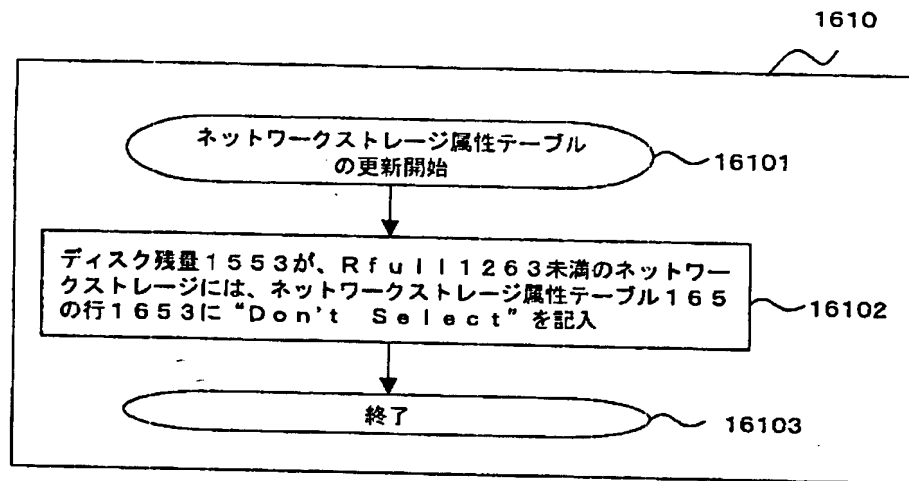
【図10】

図10

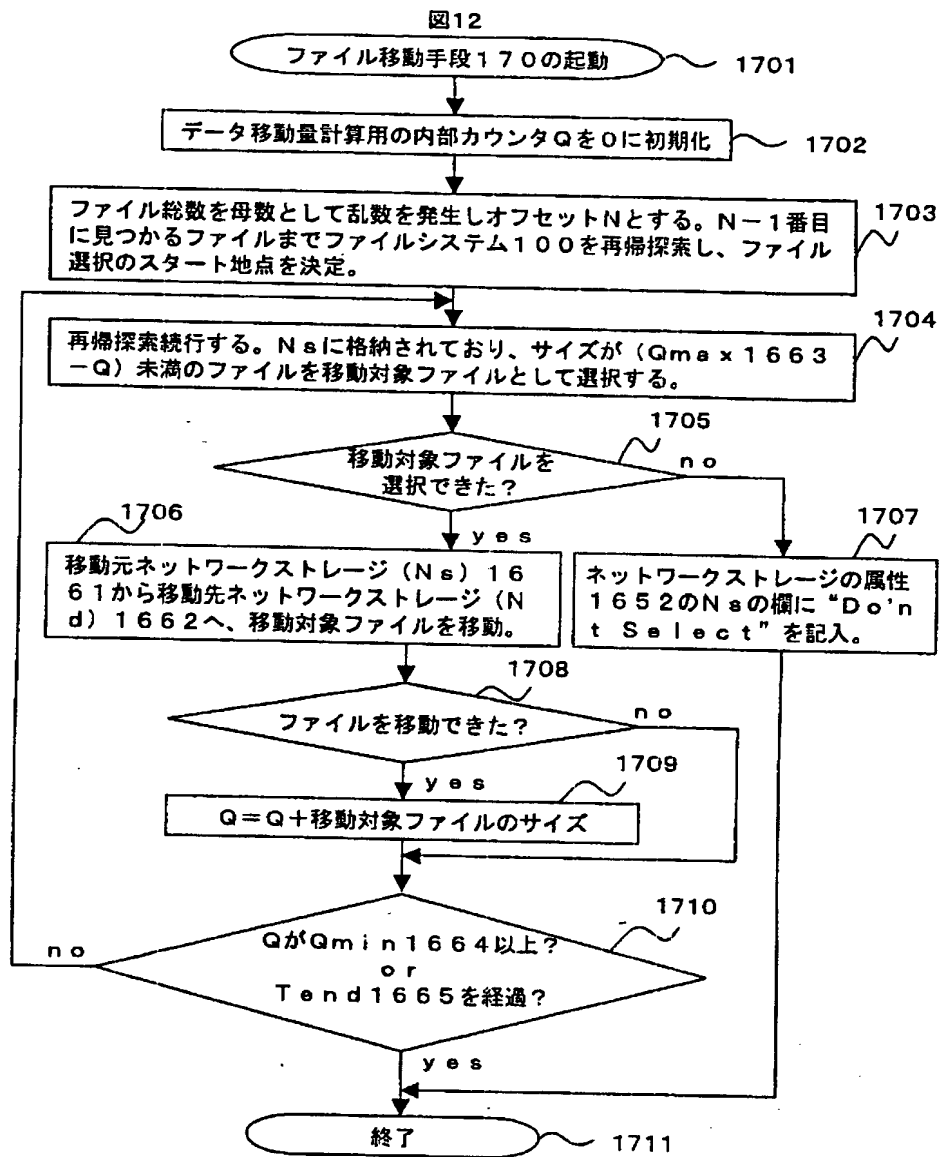


【図 11】

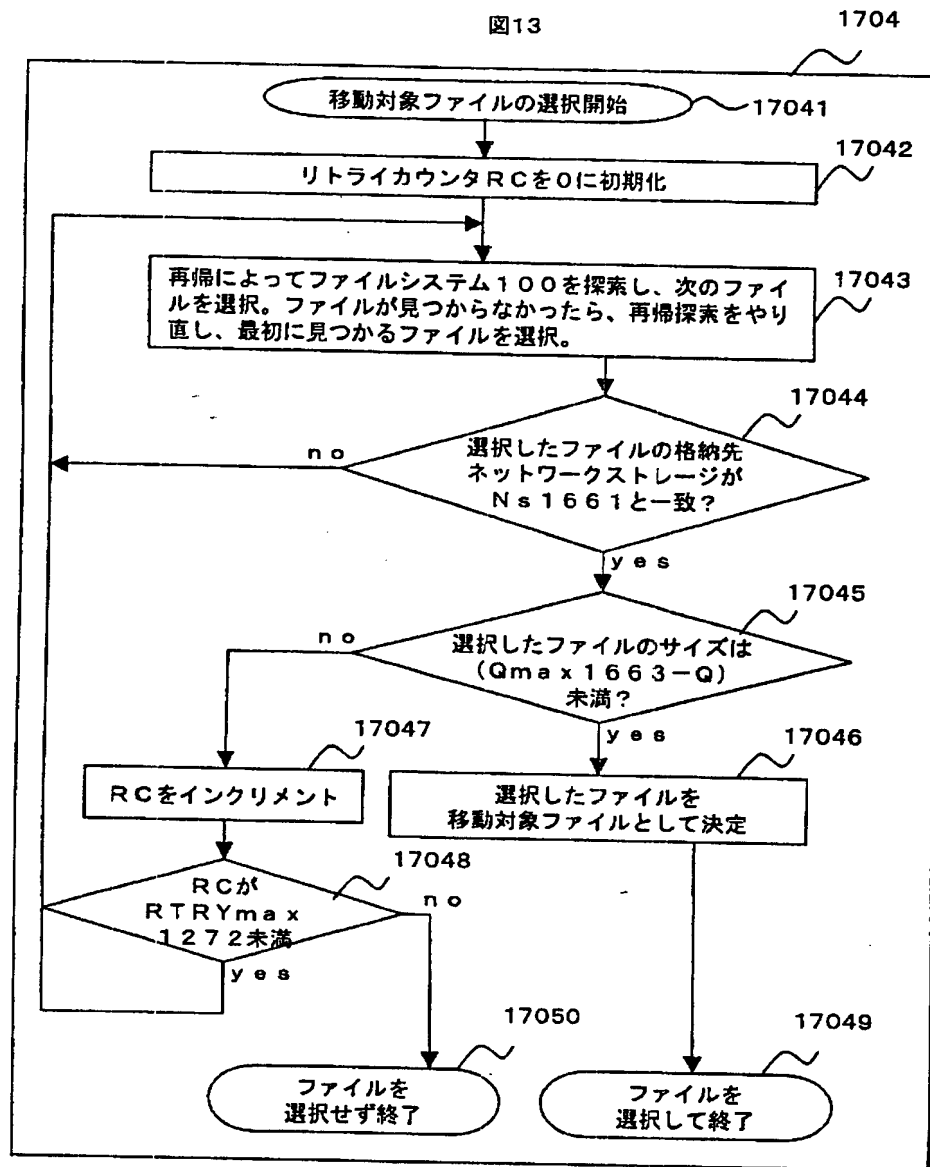
図11



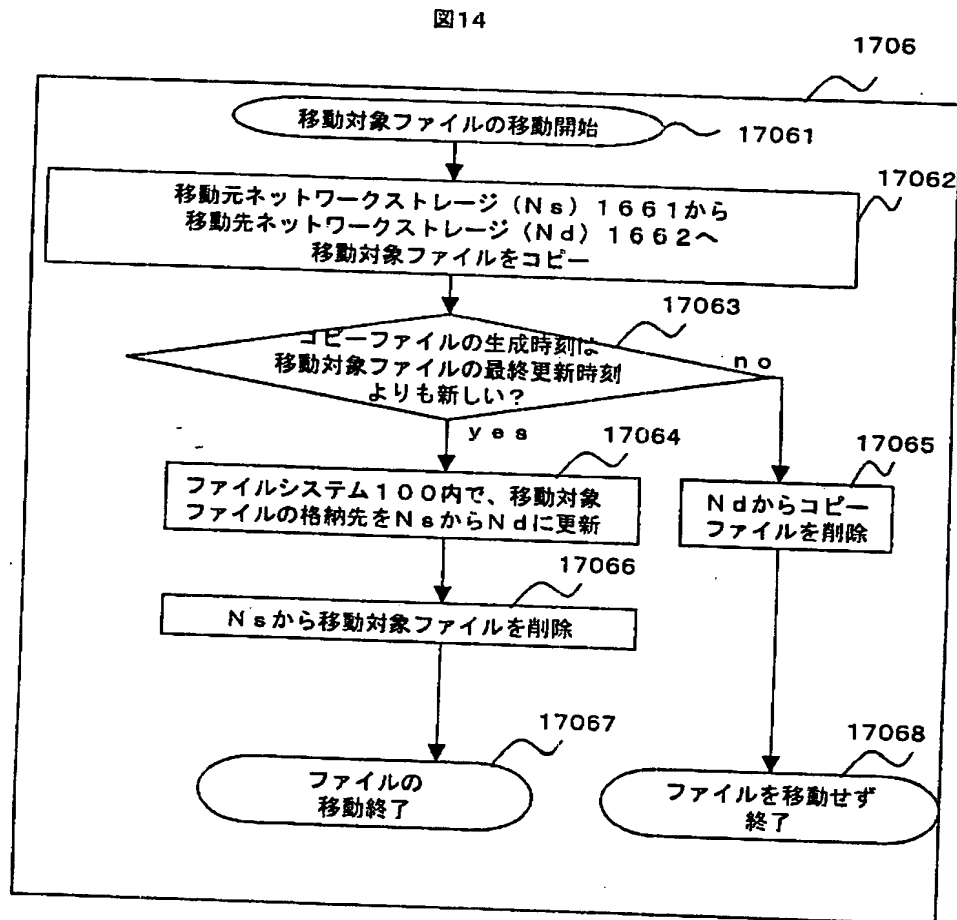
【図 12】



【図13】

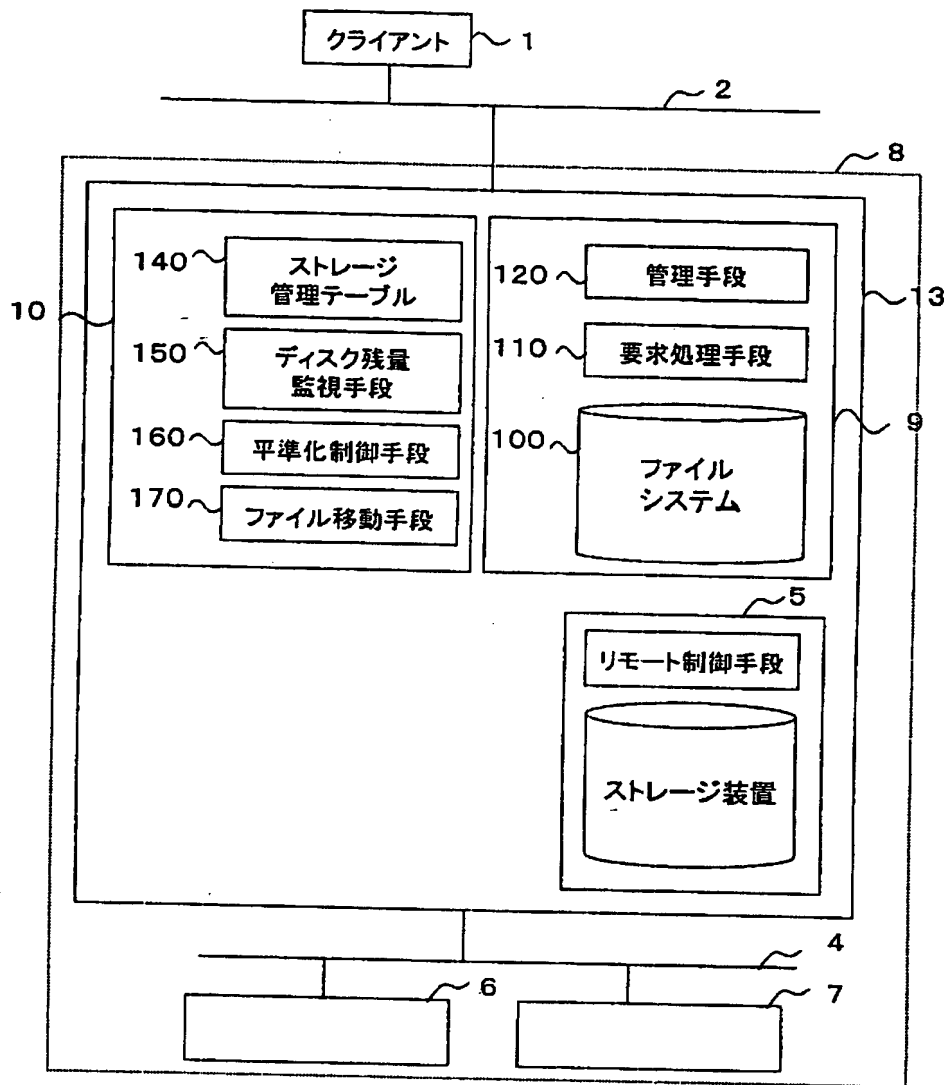


【図 14】



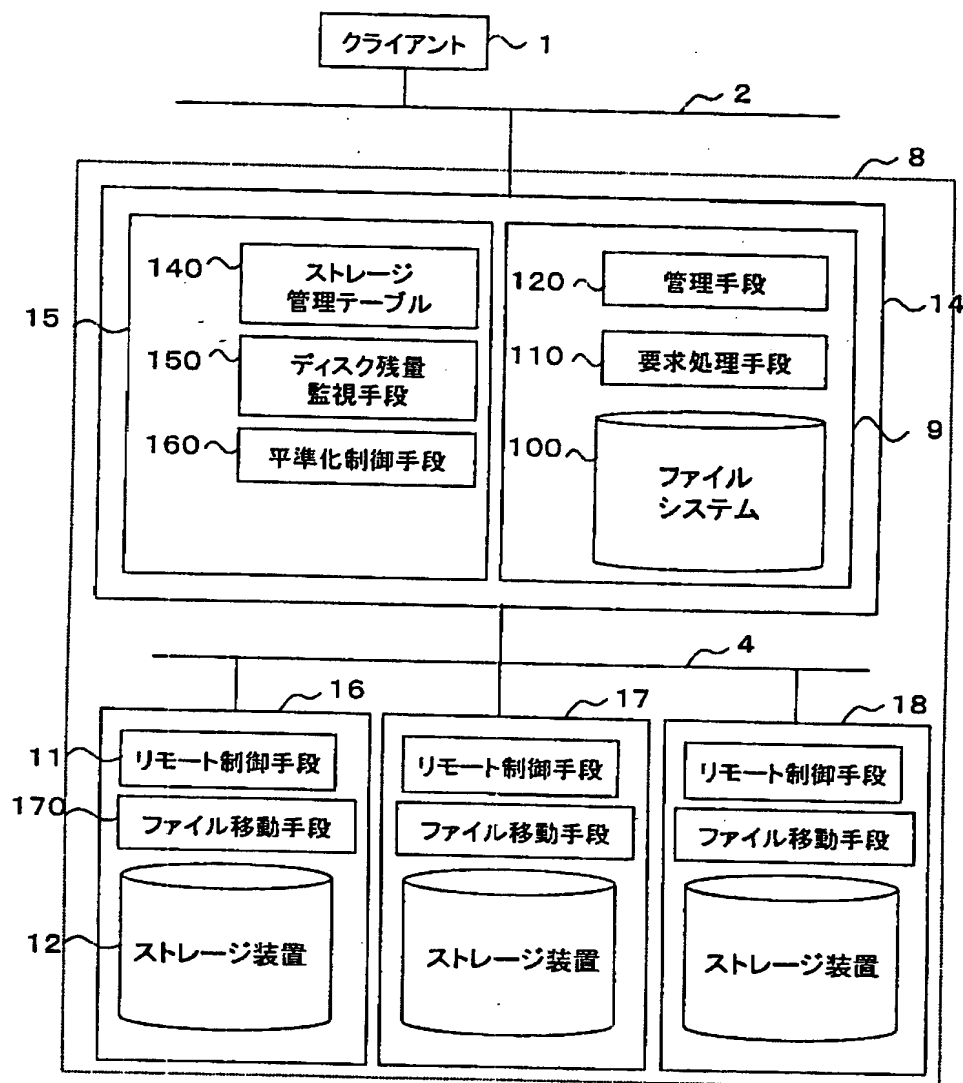
【図 15】

図15



【図 16】

図 16



【書類名】 要約書

【要約】

【課題】 各ネットワークストレージのディスク容量が不均一な仮想一元化ネットワークストレージシステムにて、ユーザが使用可能な領域、ファイルへの大量書込みに対するマージンを最大化したディスク使用量の平準化方法を提供し、また、クライアントに対するファイル共有サービスを止めることなく上記の平準化処理を行なう。

【解決手段】 各ネットワークストレージのディスク残量を調査し、ネットワークストレージ間でファイルを移動して、ディスク残量の平準化を行なう。また、クライアントから移動中のファイルに対して書込みを行なったかどうかを調査し、書込みがあった場合はファイルの移動を中止し、別のファイルを移動する。

【効果】 定常的な不均衡が解消される上に、ファイルへの大量書込みに対するマージンが最大化されるため、システム全体のディスク容量が有効に利用可能となる。また、ディスク残量の平準化処理中にもクライアントに対するファイル共有サービスを継続することが可能となる。

【選択図】 図 8

認定・付加情報

特許出願の番号	特願 2 0 0 2 - 2 5 2 3 4 5
受付番号	5 0 2 0 1 2 9 2 5 4 1
書類名	特許願
担当官	第七担当上席 0 0 9 6
作成日	平成 1 4 年 9 月 2 日

< 認定情報・付加情報 >

【提出日】 平成14年 8月30日

次頁無

特願 2 0 0 2 - 2 5 2 3 4 5

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 5 1 0 8]

1. 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台 4 丁目 6 番地

氏 名

株式会社日立製作所